

Szabadalmak igénypontgráfjának automatikus előállítása és hibaelemzése

Kiss Márton¹, Vincze Veronika¹, Nagy Ágoston¹, Alexin Zoltán²

¹ Szegedi Tudományegyetem, Informatikai Tanszékcsoport
6720 Szeged, Árpád tér 2.
{mkiss, vinczev, nagyagoston}@inf.u-szeged.hu
² Szegedi Tudományegyetem, Szoftverfejlesztés Tanszék
H-6720 Szeged, Árpád tér 2.
alexin@inf.u-szeged.hu

Kivonat: Az alább ismertetett kutatásaink középpontjában az angol nyelvű szabadalmak igénypontjai állnak. A szabadalmak a részletes leíráson túl, az igénypontokban szabatosan foglalják össze a kért szabadalom lényegét, azt, hogy a védelem pontosan mire terjedjen ki. Egy szabadalom igénypontjai között vannak kitüntetett főigénypontok és aligénypontok, az aligénypontok főigénypontra és egymásra hivatkozhatnak. Ez az igénypontstruktúra minden esetben egy gráfot alkot. Nyelvtechnológiai eszközökkel előállítottuk az igénypontgráfot. Az előállított gráfot ábrázoltuk, hogy megkönnyítsük a hibadetektáláshoz szükséges szabályrendszer kialakítását, valamint a további kutatásokat. Mivel tanuló- vagy referenciakorpusz nem állt rendelkezésünkre így másik rendszerrel hasonlítottuk össze eredményeinket. A gráfok elemzése közben kialakítottunk egy szabályrendszert, amely megsértése legtöbbször rossz hivatkozásra, a főigénypont hiányára vagy más hibára utalt. A szabályrendszer segítségével a főigénypontok detektálására is lehetőség nyílik. A módszerrel az Amerikai Szabadalmi Hivatal által elfogadott és nyilvánosan elérhető szabadalmak között kerestünk és találtunk hibásakat.

1 A szabadalmak felépítése

A szabadalmak egységes szerkezettel bírnak [1]. A főigénypont mindig azzal kezdődik, hogy milyen kategóriába tartozik a levédetni kívánt szabadalom, például módszer, eljárás, eszköz, összetétel. Eztán következik ezek kifejtése: milyen lépésből/anyagokból áll a főigénypont elején említett dolog, és ezeket az alpontokat rekurzívan továbbfejtik az úgynevezett aligénypontokban. Fontos megjegyezni, hogy egy szabadalomnak speciális esetben több főigénypontja is lehet. A mi kutatásaink csak a főigénypont szerkezetére és az egymásra való hivatkozásaikra korlátozódtak.

2 Az igénypontgráf előállítás

Miért volt szükségünk az igénypontgráf előállítására, hiszen már van működő rendszer [2, 3], mely ezt a problémát megoldja? - tehetnénk fel joggal a kérdést. Sajnos az a rendszer, melyet mi találtunk (pattools.com/claim_tree.html) csak a gráfot állítja elő, a hivatkozások típusát viszont nem adja meg. Nekünk pedig szükségünk volt erre az információra is a további kutatáshoz.

Az igénypontok közötti kapcsolatot az igénypontokban lévő, reguláris kifejezésekkel felismerhető, hivatkozások/utalások segítségével határoztuk meg. Ezen hivatkozások felhasználásával építettük fel az igénypontgráfokat. A kutatáshoz írt programokat az UIMA keretrendszerben [4, 5] írtuk.

2.1 Az igénypontgráf előállításakor használt hivatkozástípusok

Kutatásunk során 997 db A24F alosztályba tartozó szabadalmat vizsgáltunk. A szabadalmak igénypontszekciói összesen 16812 darab igénypontot tartalmaztak. Az alábbi táblázat tartalmazza, hogy milyen hivatkozástípusokat különböztettünk meg és ezeknek milyen volt az eloszlásuk az általunk vizsgált 997 szabadalomban.

1. táblázat: A hivatkozástípusok megoszlása az általunk vizsgált 997 szabadalom esetében.

Hivatkozástípus	Előfordulás
root/nem hivatkozik	2 787
in claim #	3 277
of claim #	9 102
according to #	2 833
összes hivatkozás	17 999

2.2 A előállított igénypontgráfok ellenőrzése

Nem állt rendelkezésünkre referenciakorpusz, így egy meglévő rendszerrel hasonlítottuk össze eredményeinket. A pattools.com/claim_tree.html címen elérhető rendszer által generált gráfokkal vetettük össze a mi kimeneteinket. Így kézi ellenőrzésre csak akkor volt szükség, amikor különbséget fedeztünk föl a két kimenet között.

3 Főigénypontok meghatározása az igénypontgráf felhasználásával

Későbbi kutatási témát jelenthet, hogy a gráfokat felhasználva automatikusan detektálhatjuk a főigénypontokat. Erre nagy szükségünk lesz, mert a K+F projektünkben a későbbi szemantikus elemzés kiindulópontjai minden esetben a főigénypontok.

4 Hibaelemzéshez szükséges szabályrendszer kialakítása

Az igénypontgráf megalkotása után a kapott gráfokat elemezve 3 fő hibatípust tudunk megkülönböztetni: 1) saját magára hivatkozik az igénypont, 2) a hivatkozott igénypont nem létezik, 3) ugyanaz két igénypont száma. Valamint felderítettünk lehetséges hibákat is, melyek nem minden esetben bizonyultak hibának, így ezek jelzése után kézi ellenőrzéssel kellett eldönteni, hogy valós volt-e a jelzés. Ilyen volt például, ha egy igénypont az utána következő igénypontra hivatkozik, vagy ha a hivatkozott főigénypont és a hivatkozó igénypont között van főigénypont.

A vizsgált 997 Amerikai Szabadalmi Hivatal által elfogadott szabadalomban az alábbi táblázatban felsorolt hibákat derítettük föl.

2. táblázat: A szabadalmakban felderített hibák.

Hibatípus	Előfordulás
Saját magára hivatkozik az igénypont	6
A hivatkozott igénypont nem létezik	2
Ugyanaz két igénypont száma	4
összes detektált hiba	12

5 Összefoglalás

Módszerünk más rendszerrel való összehasonlítása és a felderített hibák elemzése azt mutatja, hogy indirekt módon bizonyítható, hogy a rendszer kis hibával működik, ezért a későbbiekben jól használható szabadalmak igénypontgráfjainak előállítására. Az igénypontgráfot felhasználva több hibatípus kiszűrhető és megállapíthatóak a főigénypontok is.

Köszönetnyilvánítás

A kutatás – részben – a MASZEKER kódnevű projekt keretében a Nemzeti Fejlesztési Ügynökség, illetve a TÁMOP-4.2.1/B-09/1/KONV-2010-0005 jelű projekt keretében az Európai Unió támogatásával, az Európai Regionális Fejlesztési Alap és az Európai Szociális Alap társfinanszírozásával valósult meg.

Bibliográfia

1. Vincze V., Nagy Á., Klausz Á., Almási A., Kiss M.: Nyelvészeti problémák a szabadalmak feldolgozásában. In: Tanács A., Vincze V. (szerk.): VII. Magyar Számítógépes Nyelvészeti Konferencia. Szegedi Tudományegyetem, Szeged (2010) 168–179
2. Milton, H. W.: Method for preparing a claim tree in the preparation of a patent application. In: Patent. Bloomfield Hills, MI, US (2008)

3. Kahn, M. R.: Patent claim visualization system and method. In: Patent, Westampton, NJ, US (2009)
4. Osenga, K.: Linguistics and patent claim construction. Rutgers Law Journal Vol. 38, No. 61 (2006) 61–108
5. D. Ferrucci, A. Lally: UIMA by Example. IBM Systems Journal 43, No. 3 (2004) 455–475
6. D. Ferrucci, A. Lally: UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. In: Journal of Natural Language Engineering. (2004) 327–348